# Critical Incident Summary (CIS)

## DFO – Multiple Applications

Incident Number: 10153453
Date: 2019-08-22
Version: version 1.0
Classification: Protected A

| | | |
|---|---|---|
| Impacted Critical Business Applications or Services (CBAS) | | Canadian Hydrographic Service Directory (CHSDir2)<br>Client Service Digital Data Portal (CSDDP)<br>Common Operating Picture (COP)<br>Electronic Knowledge Management Environment (EKME)<br>Email<br>National Electronic Access Solution (eAccess)<br>National Online Licensing System (NOLS)<br>Traverse |
| Impacted Departments | | Fisheries and Oceans Canada (DFO) |
| Incident # | | 10153453 |
| Service Management Tool | | Enterprise Control Desk (ECD) |
| Directorate | | Windows and Virtualization Management |
| Service | | Midrange - Windows and Virtualization Management |
| Responsibility Code (RC) | | SSC |
| Critical Incident Duration | Actual Start Date & Time | 2019-08-12 11:30 (ET) |
| | Reported Start Date & Time | 2019-08-12 12:09 (ET) |
| | Resolution Date & Time | 2019-08-14 15:00 (ET) |
| | Business Outage Duration | 51 hours 30 minutes |
| Areas Consulted (The following areas were consulted to review the CIS prior to publication) | | DCS - Linux/Unix Management<br>DCS - Storage, Backup & Archive<br>DCS - Windows and Virtualization Management<br>NSDS - Email and Workplace Technology Services<br>NSDS - Infrastructure Security Operations<br>SDM - Client Executive Directorates<br>SDM - Service Management |
| CIS Author | | Samer El-Kadri |
| CIS Publish Date | | 2019-08-22 |
| CIS version | | 1.0 |

# Issue

All DFO users were unable to access multiple applications and users in Ontario, Québec and Arctic region were unable to connect to the network nor send or receive email due to failure of the VMware ESXi host server cluster.

# Background

The business hours for Email services and National Online Licensing System (NOLS) are 24/7 while support hours of their related storage unit is Monday to Friday from 06:00 until 20:30 ET. The remaining DFO impacted critical applications are supported Monday to Friday from 06:00 ET until 20:30 ET.

DFO's Data Centre (DC) located at [redacted] hosts multiple production ESXi host server clusters. The impacted applications were part of the [redacted] environment, which is a cluster of 6 VMware ESXi host servers hosting 163 Virtual Machines (VMs) that connect through two fibre channel Storage Area Network (SAN) Brocade switches to a Hitachi storage array.

Purple Screen of Death (PSOD) is a diagnostic screen with white type on a purple background that is displayed when the Operating System (OS) of an ESX/ESXi host server experiences a critical error, becomes inoperative and terminates any virtual machines that are running.

A Logical Unit Number (LUN) is a data storage unit that constitutes a configured virtual set of disk drives that is presentable to a host and mounted as a volume within the VM cluster. The storage array within the [redacted] environment is segmented into 11 LUNs of 4 Terabytes each with a combined total of 44 Terabytes.

A datastore is a reference by vendor VMware for a storage area within a data storage unit (LUN) that VMs can reside on.

# Description of Incident

August 12, 2019

At 12:09, the DFO Service Desk reported to the SSC Enterprise Service Desk (ESD) that users were unable to access the Internet. At 13:11, the SSC ESD informed the SSC Incident Coordination Team (ICT) of the High priority incident. SSC LAN support was immediately engaged by SSC ICT to investigate. At 13:52, the SSC ESD notified the SSC ICT that the SSC Client Executive (CE) for DFO requested the priority be raised to Critical as multiple critical applications, including Email, were impacted. SSC ICT immediately contacted SSC Windows and Virtualization support, who advised that multiple VMs were affected by the failure of all 6 of [redacted] s ESXi host servers, which had experienced Purple Screen of Death (PSOD), causing multiple critical applications to be unavailable. SSC Storage support was engaged and found no hardware faults on the storage array during their investigation as it was online and serving production data without any alerts. SSC Storage support investigated the logs of [redacted] s SAN switches and found no errors, but noted log entries indicating that the [redacted] 6 ESXi host servers hosting multiple VMs were now offline. Subsequently, SSC Windows and Virtualization support engaged the vendor (VMware) to join an ongoing technical conference call

to assist in the investigation. Log files were shared and analysed amongst the investigating parties in a combined effort attempting to identify the root cause.

At 17:30, based on analysis of the logs, VMware suspected that there might be drive corruption on the storage array and requested SSC Storage support engage the vendor (Hitachi) to assist in the investigation, who ran complete hardware diagnostics. At 19:21, Hitachi had joined the technical conference call, and reported that they found no issues from their diagnostics. Meanwhile, SSC Storage support continued to work with SSC Windows and Virtualization support and both vendors on narrowing down the list of potential problematic LUNs. By individually removing and reallocating the LUNs to one of the impacted ESXi host servers that they had managed to successfully bring back online. This method was being used to identify and isolate the LUNs that were causing [redacted] ESXi host servers to fail. By 20:24, SSC Windows and Virtualization support were able to restore 19 of the impacted VMs on an ESXi host server that it had successfully brought back online, however VMs were not considered stable.

By 21:55, a second ESXi host server was also successfully brought back online and it was recommended by SSC Windows and Virtualization support to be utilized instead of the first ESXi host server that was now hosting 19 functioning VMs. This was done as a precaution in order to safeguard the 19 VMs from failing again, while testing other LUNs on the second ESXi host server in case it was presented with the suspected faulty LUNs. At 22:23, the second ESXi host server, which was being utilized to isolate the faulty disk drive experienced a PSOD once it was presented with the same five LUNs that were identified earlier on as not corrupted. The initial hypothesis of suspected corrupted LUNs as the root cause of the incident was discounted after testing determined it to be inconclusive.

As part of investigative efforts, SSC Windows and Virtualization support reported at 23:18 that they had updated the firmware of the fibre channel host bus adapters, which were not at a supported firmware level version, and that the correct firmware was installed on two of the hosts, which still experienced PSOD.

August 13, 2019

At 00:50, the technical conference call ended and was scheduled to reconvene at 08:30 to continue with the investigation given that it was outside support hours for SSC Storage support and resources were unavailable at that time.

At 08:30, the technical conference call was reconvened. VMware confirmed that based on their analysis of the PSOD logs, they had found a common error across all 6 ESXi host servers as they were reviewing the specifics in detail while planning the next steps forward. At 08:52, SSC Email support reported the restoration of Email services including mobile devices after SSC Enterprise Mobile Device Management (EMDM) support restarted the Certmode servers for DFO. This was a confirmation that first recovered ESXi host server was hosting the VMs that enabled access to Email services.

DFO requested the focus to be on bringing up critical applications; however, SSC support groups did not have a list of which applications were hosted on which of the 163 VMs. DFO also did not have the information readily available.

By 09:35, the SSC Windows and Virtualization support and VMware had reverted one of the ESXi host servers and its network driver to a previous known stable version in order eliminate the possibility of root cause being related to version upgrades, however, the PSOD error still persisted. As part of troubleshooting steps, VMware requested that SSC Storage verify the physical connection to the [redacted] ESXi host server cluster. A local SSC Storage support was dispatched to the site and confirmed that the physical connections were intact.

By 11:13, SSC Storage requested that VMware escalate the incident to Level 2 support once the analysis of the PSOD log files were completed. At this time, SSC Windows and Virtualization support decided to begin migrating VMs off an operable ESXi host server in alternate VM cluster environment [redacted] and deploy the ESXi host server to [redacted] environment in an attempt to restore critical DFO services. At 11:54, VMware Level 2 support completed their preliminary analysis of the PSOD log files and were unable to locate any errors that would point towards a root cause. VMware escalated the incident to Level 3 support, who joined the ongoing technical conference call at 12:06.

By 12:41, Hitachi had confirmed no errors were detected on the disk drives within the storage array. SSC Windows and Virtualization support also engaged a second storage vendor (EMC), who confirmed successful health checks on the SAN switches with no errors being detected. At 13:55, the newly deployed ESXi host server from [redacted] environment experienced a PSOD error as the other servers once it became a member of the [redacted] environment; SSC Storage support and VMware shared the PSOD log files with Hitachi for analysis.

At 14:24, SSC Windows and Virtualization support launched a system file check verification on a datastore of one of the LUNs that VMware suspected to be faulty. The system file check verification completed at 15:33, and errors were found in the datastore. By 15:55, SSC Windows and Virtualization support had decided to launch an advanced fix procedure after their failed attempt to fix the errors using the simple fix procedure.

At 16:46, the operational ESXi host server 16 experienced PSOD error causing the 19 hosted VMs to fail, impacting access to Email services.

At 18:00, VMware invited additional resources to assist with the investigation. The advanced fix procedure earlier mentioned was no longer being considered due to the extent of the data corruption. VMware relaunched an investigation into the PSOD log files starting from the onset of the incident, while SSC support resources discussed the possibility of restoring the VMs from the most recent backup.

At 18:57, it was decided to proceed with an option of creating 5 builds of the non-configured VMs required to restore Email and NOLS once SSC Storage support allocates 44 Terabytes of storage space to be used in the restoration of the affected [redacted] VM cluster. By 19:55, SSC Storage support successfully allocated the required 44 Terabytes of storage space, and SSC Windows and Virtualization support initiated the process of creating the 5 builds of the non-configured VMs.

By 20:23, VMware noted that the logs were showing an increase just before first ESXI host servers in the Prod 2 environment failed when it encountered a PSOD and provided a list of most likely LUNs to have corruption. At the request of VMware, SSC Storage support unallocated the storage capacity from all the [redacted] ESXi host servers, which constituted 11 LUNs in total, and allocated only one LUN to an isolated ESXi host server to allow VMware to scan it using a specialized tool that checks metadata consistency of the file system and to verify it for any errors. VMware's scan of the LUN revealed errors on some LUNs, however there were also stale locks on the data storage unit's (LUN) datastore, which can prevent the ESXi host server from accessing the VMs resulting in the ESXi host server to experience a PSOD.

At 21:30, VMware recommended a technical procedure to eliminate the stale locks by cloning the LUN, which would allow the ESXi host server to detect it and allow it to clear the stale locks, thereby restoring access to the datastore and its VMs.

In parallel, at 22:50, SSC Windows and Virtualization support had completed 3 of the 5 builds of the non-configured VMs required for the service restoration of Email and National Online Licensing System (NOLS).

August 14, 2019

By 00:20, SSC Windows and Virtualization support was patching and configuring 3 of the newly built VMs, while, SSC Storage support continued engaging Hitachi in the procedure to eliminate the stale locks. Once completed, a scan on the cloned LUN showed data corruption as stale locks persisted, where they were cleared. SSC Storage support then unallocated the cloned LUN from the ESXi host server and allocated it to another ESXi host server with access to the newly allocated 44 Terabyte storage.

At 01:30, SSC Windows and Virtualization support and VMware agreed to use an alternate suggested method of clearing the locks on subsequent LUNs, which did not involve cloning the LUNs and was much less time consuming. The new method was used on the remaining 10 LUNs, where the datastores were successfully scanned and cleared of any data corruption or locks.

By 03:25, SSC Windows and Virtualization support had completed the patching and configuration process for 4 of the 5 VM builds while SSC Windows and Virtualization support and VMware successfully recovered access to more than half of the LUNs and continued to focus their recovery and restoration efforts on the remaining LUNs.

At 04:25, SSC Windows and Virtualization support and VMware were able to successfully recover all of the LUNs without any data loss. By 05:30, SSC Windows and Virtualization support successfully recovered and mounted all the VMs on the ESXi host servers as they powered on 9 of the 163 VMs. At this point, it was decided by SSC Windows and Virtualization support to halt and abandon the recovery and restoration option involving patching and configuration process of the 5 VM builds due to the success of the alternate option of recovering the VMs.

At 06:40, active directory profile issues were encountered, where SSC ICT engaged SSC Active Directory Services support to investigate.

By 08:35, The SSC Directory Services confirmed all Active Directory servers were online and replicating correctly. SSC Linux/UNIX support was engaged by SSC ICT to verify the state of their related VM infrastructure as part of the VMs that were being brought online.

At 09:05, SSC Windows support confirmed successful health checks on DHCP servers. At 09:30, NOLS and eAccess were confirmed by the partner to be back online and operating normally.

By 09:55, SSC Windows and Virtualization support continued to sequentially bring back online the remaining VMs to ensure dependencies were resolving correctly, where CHSDir2 and Traverse were confirmed by the partner to be back online at 10:30 and 11:15 respectively, and were operating within normal parameters.

By 11:40, 40 servers had been brought back online, and remained stable yet performance degradation was observed on the Email VM only while other VMs were performing as expected. SSC Legacy Email support had already been actively investigating the performance issues and was in contact vendor (Microsoft).

Confirmation of CSDDP, COP and EKME was received from the partner at 13:30. By 14:00, 100 VMs had been brought back online and remained stable. SSC Windows and Virtualization support continued to work on bringing the remaining 63 VMs back online and continued working with SSC Legacy Email support and Microsoft to identify the Email performance issues.

By 14:45, Email services were restored as it was identified that Active Directory permission issues were fixed for the Exchange servers. Email for BlackBerry and Android mobile devices were restored with the exception of iPhone mobile devices that still were experiencing issues with sending and receiving emails. At 15:00, SSC EMDM support successfully restarted the Certmode servers for DFO, which resumed communications between the exchange server and iPhone mobile devices, restoring their Email services. As requested by the partner, confirmation of Email service restoration was only possible the next day.

August 15, 2019

Confirmation of Email service restoration was received from the partner at 08:05.

# Business Impact

**NOLS**

DFO employees were unable to track payment for fishing licenses due to NOLS being unavailable. Canadians were unable to obtain fishing licenses as licensing offices are no longer staffed to process paper documents as the entire process is now online with NOLS as the main user gateway.

**CSDDP**

DFO Employees were unable to allow client distributors to see via the CSDDP portal the charts, products and hydrographic data that are available for license as updates to purchased navigational products are acquired solely through the CSDDP portal.

**COP**

Due to the unavailability of COP, DFO employees had difficulty collecting and displaying vessel position information across several ports. An alternate method was available to the employees, which involved the use of other systems, but it made the analysis slow and inefficient. This decreased efficiency of regional operations centres presented a risk of potentially causing shipping delays or safety concerns as it is a critical component of Maritime Search and Rescue activities and information is used in rescue operations and helps to save lives that were at risk.

**EKME**

DFO employees did not have access to many information records of business value stored in DFO's official repository, EKME.

**Traverse**

DFO employees were unable to provide chart sales as Traverse is used to issue invoices, maintain client accounts, payment history, and to issue revenue reports.

**CHSDir2**

DFO Hydrographers was unable to store information related to hydrographic surveys, new constructions navigational pathway and any information relevant in creating navigational charts, an atlas, etc. Cartographers were unable to update a product such as a navigational chart as the Updating Group uses the system to check the new information added by the hydrographers and to decide if a notice to mariners is needed.

**Email**

Employees in Ontario, Québec and Arctic region were unable unable to communicate effectively.

# Post Incident Activities & Recommendations

All DFO users were unable to access multiple applications and up to 5,000 users in Ontario, Québec and Arctic region were unable to connect to the network nor send or receive email due to failure of the ESXI host server cluster, which had encountered stale locks on the LUNs, preventing them from accessing the VMs found within the LUNs datastores.

- SSC Storage support and SSC Windows and Virtualization support to work with vendors Hitachi and VMware to determine root cause of the encountered stale locks on the LUNs and to identify and implement improvements to prevent recurrence.

The business hours for Email services and National Online Licensing System (NOLS) are 24/7 while support hours of their related storage unit is Monday to Friday from 06:00 until 20:30 ET.

- SSC CE for DFO to work with the partner and SSC Storage support to review and confirm the support hours for storage support for DFO.

During the incident, DFO requested priority be given to particular critical applications to be brought up once ESXi servers and their associated VMs stabilize. Neither DFO nor SSC Windows and Virtualization support have a list outlining which VMs host them.

- SSC CE for DFO to work with the partner and SSC Windows and Virtualization support to review and confirm the list of critical applications with respect to hosting VMs.

This report formally engages Problem Management. A problem record is created for all critical priority incidents to determine the requirement for root cause investigation. Inquiries into any further investigation should be directed to the following mailbox: SSC.probleminfo-infoprobleme.SPC@canada.ca

**The sensitivity of the information entered in this document is limited to information at the Protected A level, as a maximum. This document is to be handled in accordance with the Policy on Government Security with access restricted to authorized individuals whose duties require such access on a "need-to-know" basis.**